BEST AVAILABLE COPY

PCT/EP 00 / 13206

09/914240

Europäisches Patentamt

European Patent Office

Office européen des brevets

REC'D 12 MAR 2001

WIPO PCT

EP 00/13206

4

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. **Patent application No.** **Demande de brevet n°**

99403308.2

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(A) OR (B)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

Hatten

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 14/11/00
LA HAYE, LE

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 99403308.2
Demande n°:

Anmeldetag:
Date of filing: 28/12/99
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
SNR scalable video coding using hierarchical meshes and triangle-based matching pursuit

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:
State:
Pays:

Tag:
Date:
Date:

Aktenzeichen:
File no.
Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

SNR SCALABLE VIDEO CODING USING HIERARCHICAL MESHES AND TRIANGLE-BASED MATCHING PURSUIT

Authors: Vincent Bottreau, Marion Bénetière, Béatrice Pesquet-Popescu

BACKGROUND

Scalability is an important research topic in video compression that has attracted considerable attention recently. Scalability is the expected functionality to address the ever growing constraints of video transmission over heterogeneous networks (bandwidth, error rate...) in terms of varying receiver capabilities and demands (CPU, display size, application). Scalability allows a progressive transmission of information (in layers or not) in order to provide a quality level of the reconstructed video sequence that is proportional to the amount of information that is taken of the bitstream.

Although they have not been initially designed to address these issues, current standards tried to upgrade their video coding schemes in order to include this functionality. In quality or SNR scalable compression schemes, temporal and spatial resolutions are kept the same, but the image quality is intended to vary depending on how much of the bitstream is decoded. In practice, most standards provide SNR scalability by means of a layered structure without giving up the classical single-scale scheme. The base layer (BL) is generally highly and efficiently compressed by a hybrid predictive encoding loop. The enhancement layer (EL) improves the quality of the compressed video signal by encoding the residual error (or prediction error), which is the difference between the original image and the reconstructed image. In MPEG-4 version 4 for instance, the EL uses DCT bit-planes to reencode this residual error [1]. However, the resulting scalability is suboptimal for two main reasons:

- It is only based on an additional encoding of the prediction error and does not involve any refinement of the motion estimation and compensation processes, whereas a global approach that refines the whole scheme may achieve a better reconstruction.
- It employs coding techniques like DCT that are not intrinsically designed to provide a progressive information transmission.

Therefore, normalization committee experts are looking toward new breakthroughs for efficient scalable video coding. Hierarchical strategies appear to be the most promising candidates [2]. The main idea is to design schemes that provide a generalized hierarchical representation of the information, which naturally opens the way to scalability. Schematically, a simple hierarchical video coding scheme may be composed of several levels, each of which delivering a better-reconstructed image by means of a global refinement process. For instance, the hierarchy may use a pyramid composed of several image resolutions. Among these research axes, multiresolution techniques such as subband decomposition are obviously getting ahead in new standards (H26L, JPEG-2000).

In parallel, hierarchical hybrid predictive coding schemes using non-block image representations like triangular meshes constitute also an interesting alternative and have given rise to more and more research efforts these last years. In this latter approach, there is a need for new prediction error coding techniques. Meshes are actually well adapted to prediction error coding since they efficiently make the distinction between smooth regions and contours, well-compensated areas and occlusion regions. However, existing mesh-based systems generally encode the prediction error in a traditional way (block-based DCT for example), that is by treating the error image as a whole picture without using the mesh employed during the motion estimation and compensation stages. To address this issue, techniques such as Shape-Adaptive DCT have been adapted to the mesh structure. Nevertheless, these methods still suffer from a lack of flexibility, especially at low bit-rates, and do not provide an embedded bitstream.

In this context, the Matching Pursuit (MP) method is an attractive method. Indeed, MP is particularly well suited to the progressive texture encoding of arbitrarily shaped objects. Moreover, an intrinsic way of providing SNR scalability with MP is through the number of encoded "atoms". MP naturally achieves scalability by encoding the motion prediction error in decreasing order of energy. The procedure is iteratively applied until either the bit budget is exhausted or the distortion falls down below a prespecified threshold. The granularity of MP is the coding cost of one atom, that is approximately 20 bits.

In a previous patent proposal[1], we have included the MP prediction error coding method inside a hierarchical mesh-based video-coding scheme. We have benefited from the triangular mesh advantages concerning spatial adaptability, deformation capacity, compact and robust motion estimation even at low bit rates. This work is based on a software library developed at CNET (Centre National d'Etudes des Télécommmunications de France Telecom), mainly composed of mesh-based coding tools (mesh generator, motion estimator, etc). The mesh hierarchy of this scheme is obtained through a coarse-to-fine strategy, beginning at the first level with a coarse regular triangular mesh. The mesh refinement process locally subdivides triangles of the current level where the prediction error signal is still important after motion compensation. The new mesh is taken as the input for the following level. Coupled to a hierarchy of low-pass filtered images, this representation gives an information accuracy that increases with the level (see Figure 1).

Considering only the hierarchical feature of these tools, it appears that they do not provide scalability on their own. The reason is that motion estimation is performed at each hierarchy level between the same source images as for the first level, without taking into account the already produced images at previous levels, which reduces the coding efficiency. To achieve a scalable coding scheme, it is a necessary to start from the data information that has already been encoded in order to avoid any redundancy or overhead and to efficiently refine this information.

Based on this hierarchical triangular mesh representation, our invention provides an intrinsically SNR scalable coding scheme, which at each level jointly refines the grid (by further splitting mesh triangles), the motion estimation and compensation processes and the texture of the motion compensated image (by coding the prediction error with a MP method adapted to the mesh structure of this level). Therefore, the reconstructed image quality progressively increases from level to level. Moreover, thanks to MP characteristics, the part of the bitstream dedicated to the prediction error texture coding is embedded.

Description of the invention

The present invention improves our previous work by efficiently combining the MP algorithm with the hierarchical feature of the mesh-based structure so as to provide SNR scalability. The targeted BL and EL have been naturally associated to the different levels of the mesh hierarchy. The BL consists of the combination of the coarsest mesh, the associated motion vectors, the MP-coded atoms and the first level reconstructed image. The BL image is the first level motion compensated image whose quality has been improved by adding atoms coming from the MP encoding of the corresponding motion residual image.

A strong requirement for scalability is that the encoder only uses the information that will be available at the decoder side so as to avoid any drift problem. This constraint constitutes the real cost of scalability. Indeed, the general issue concerning scalability is the efficient combination of two information sources: the reconstructed images obtained at previous layers inside the hierarchy for image *N* and the already encoded layers of image *N-1*. Our invention addresses this issue by taking:

- the BL of the previous image as the reference image for the current image BL motion estimation,
- the current level reconstructed image as an input for the next hierarchy level.

The original hierarchy does not provide scalability because the enhancement levels take as inputs the same images as for the first level. In more details, the coarsest mesh is refined at the first level for the next ones according to the DFD energy between the BL reconstructed image and the current image *N*. Once refined, i.e. updated by splitting triangles with the highest residual energy, this mesh is used at the second level to improve the previous motion vectors. The coarsest mesh motion vectors are propagated from parent to child nodal points and are used as initial values for a new motion estimation process between the same reference and current images. The motion estimation and motion compensation processes are thus also refined. Nevertheless, this new reconstructed image can not be easily derived from the previous level reconstructed image. The reason is that they have not been obtained with the same parameters, although both images represent an approximation of the same image, that is the current image *N*. It is actually undesirable to send to the decoder too many information overheads and a fortiori to send a second time the same information, here the motion information. In the same

[1] The MP algorithm is described in details in the patent proposal n. XXX "Prediction error coding using triangle-based Matching Pursuit in a hierarchical mesh-based video coding scheme", V. Bottreau, M. Bénetière and B. Pesquet-Popescu

manner, the corresponding motion residual image (at the second level) is MP-coded to obtain the reconstructed image of the current level, the same way as for the first one. Atoms are encoded in order to improve the texture of the motion compensated image. However, atoms contained in the first level reconstructed image are in this case not used. Therefore, encoded and transmitted atoms at previous level are no longer of any use for computing the EL at the decoder side, which is not satisfactory as far as scalability is concerned.

For these reasons, so as to improve the coding efficiency of the enhancement layers, we use the previous level reconstructed image as input for the next level of the hierarchical coding scheme. The main advantages of our invention are:

- each encoded information (motion, texture, mesh, atoms...) at a certain level is intrinsically used at the following ones since enhancement levels take as inputs the previous layer components,
- a certain level really represents the enhancement of the previous ones by progressively adding refinement data (motion vectors for motion refinement and atoms for texture enhancement),
- scalability is preserved since all processed images are available at the decoder side, which prevents from having any coding drift.

Detailed description of the invention

In this section, we present our SNR scalable coding scheme. It consists of three levels as described hereinafter. Figure 2 and Figure 3 respectively illustrate a block diagram of the encoder and a block diagram of the decoder according to the invention. Level 1 corresponds to the base layer, whereas levels 2 and 3 correspond to the enhancement layers. Potentially, this scheme may be completed with more enhancement layers.

Encoder

Let us introduce the notations used in Figure 2 and Figure 3: the encoder takes as input a couple of images (reference and current images) and a mesh (the coarsest one). ε_i stands for the error residual image between the reference image N and the motion compensated image Nc_i after the i-th level. ε_i is encoded by Matching Pursuit and reconstructed by means of the encoded atoms MP_i. This reconstructed motion residual image is added to Nc_i to produce the enhanced (or reconstructed) image Nc_i', which corresponds to the current level layer image. The new error residual image ε_i' between N and Nc_i' is used to refine the current level mesh $Mesh_i$ towards mesh $Mesh_i'$, which is taken as input for the next level, $i+1$. The information concerning the mesh distortion is contained in motion vectors (MV_i), which represent the vertex displacements. Since meshes share common nodes, it is useless to completely transmit them. It is sufficient to transmit the new nodes at each level.

Error residual images ε_i' correspond to the differences between the current image N and the motion compensated images Nc_i. The operation that produces the second and third motion compensated images is nonetheless not a motion estimation strictly speaking since it is applied between two versions of image N: image N itself and the previous reconstructed image Nc_i', i.e. the motion compensated image that has been enhanced by the MP-coded atoms. As a matter of fact, this introduces a break in the motion field. If the theoretical assumptions of this method may be questionable, it is efficient in both terms of PSNR and visual results. This method allows to exploit at the same time motion and texture data that have been brought by the previous level. Therefore, our invention provides a response to the issue of SNR scalability inside hierarchical coding schemes.

Decoder

Assuming that the first original image has been encoded in intra mode and transmitted, following inter-coded images can be reconstructed at the decoder side thanks to the information related to meshes, atoms and motion vectors contained in the three layers. Figure 3 shows the way in which the three enhanced images are reconstructed at the decoder side. Once decoded, the base layer image N_{c1}' may be refined by applying motion vectors MV_2 and adding texture information contained in transmitted atoms MP_2. Moreover, the texture enhancement provided by atoms is progressive thanks to the characteristics of the Matching Pursuit method. According to the decoder complexity, the refinement process may be carried on to the following enhancement layer.

Originality

Our invention addresses the issue of SNR scalability inside a hierarchical mesh-based video-coding scheme, which naturally offers a powerful and flexible framework for scalable applications. A Matching Pursuit

prediction error coding method, specifically adapted to the triangular mesh support is used inside a hierarchical coding scheme, which has been modified so as to provide a progressive information compression.

References

[1] Y. Chen, "Summary of mini experiment on Fine Granular Video Scalability", contribution M3987 to ISO/IEC JTC1/SC29/WG11, October 1998

[2] M. Bénetière, B. Pesquet-Popescu and Y. Ramanzin, "Advanced techniques for scalable and robust video coding : a bibliographical study", LEP-Technical Report No. C 99-710, May 1999

CLAIMS :

1. In view of an SNR scalable video coding scheme, an encoding method allowing a progressive transmission of information in a base layer BL and at least one enhancement layer EL, said encoding method being based on a hierarchical triangular mesh representation to which a matching pursuit error coding step is specifically adapted.

2. An encoder for the implementation of an encoding method according to claim 1, wherein said encoder corresponds to the block diagram of Figure 2.

3. A decoding method provided for receiving images that, before their transmission and/or storage, have been previously coded, in the form of a base layer BL and at least one enhancement layer EL, by means of an encoding process that is based on a hierarchical triangular mesh representation to which a matching pursuit error coding step has been specifically adapted, said encoding process leading to the transmission and/or storage of the first original image in an intra mode excluding any prediction and the following ones in an inter mode involving motion estimation and compensation between reference and current images, wherein said decoding method comprises a step of reconstruction thanks to the information related to meshes MH, atoms MP and motion vectors MV contained in the base layer and the enhancement layer(s), said reconstruction step itself including the successive sub-steps of decoding the base layer image, refining said decoded base layer image by applying corresponding motion vectors, and adding texture information contained in transmitted atoms.

4. A decoder for the implementation of a decoding method according to claim 3, wherein said decoder corresponds to the block diagram of Figure 3.

Figures

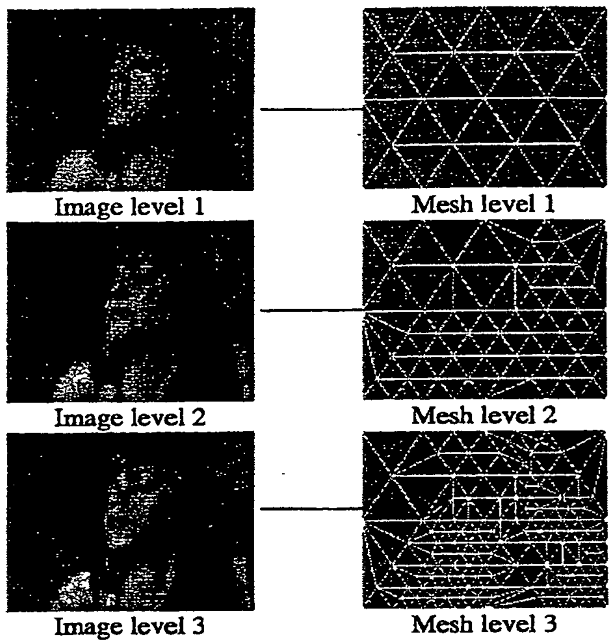

Figure 1: Image and mesh hierarchy

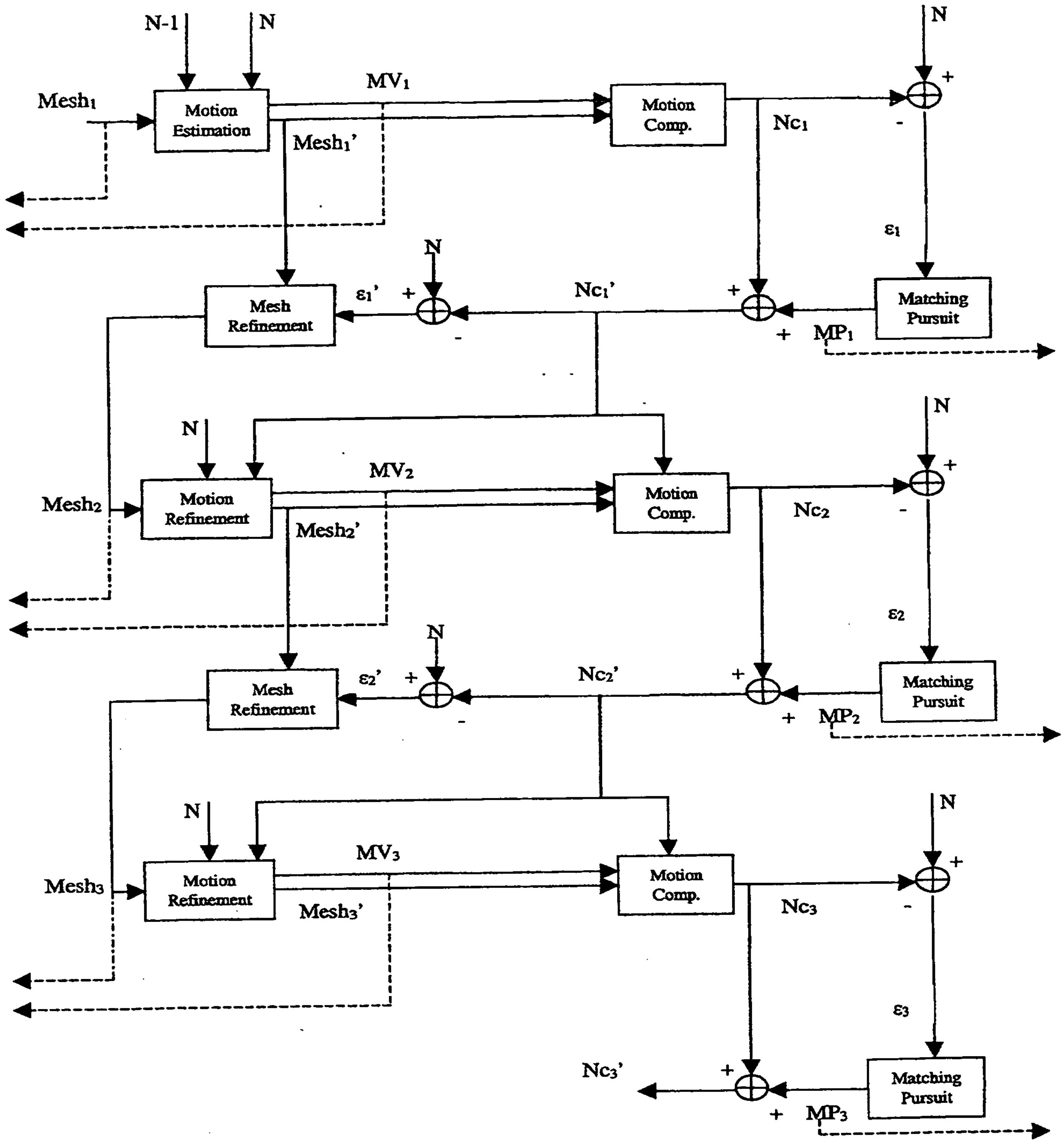

Figure 2: Block diagram of the encoder

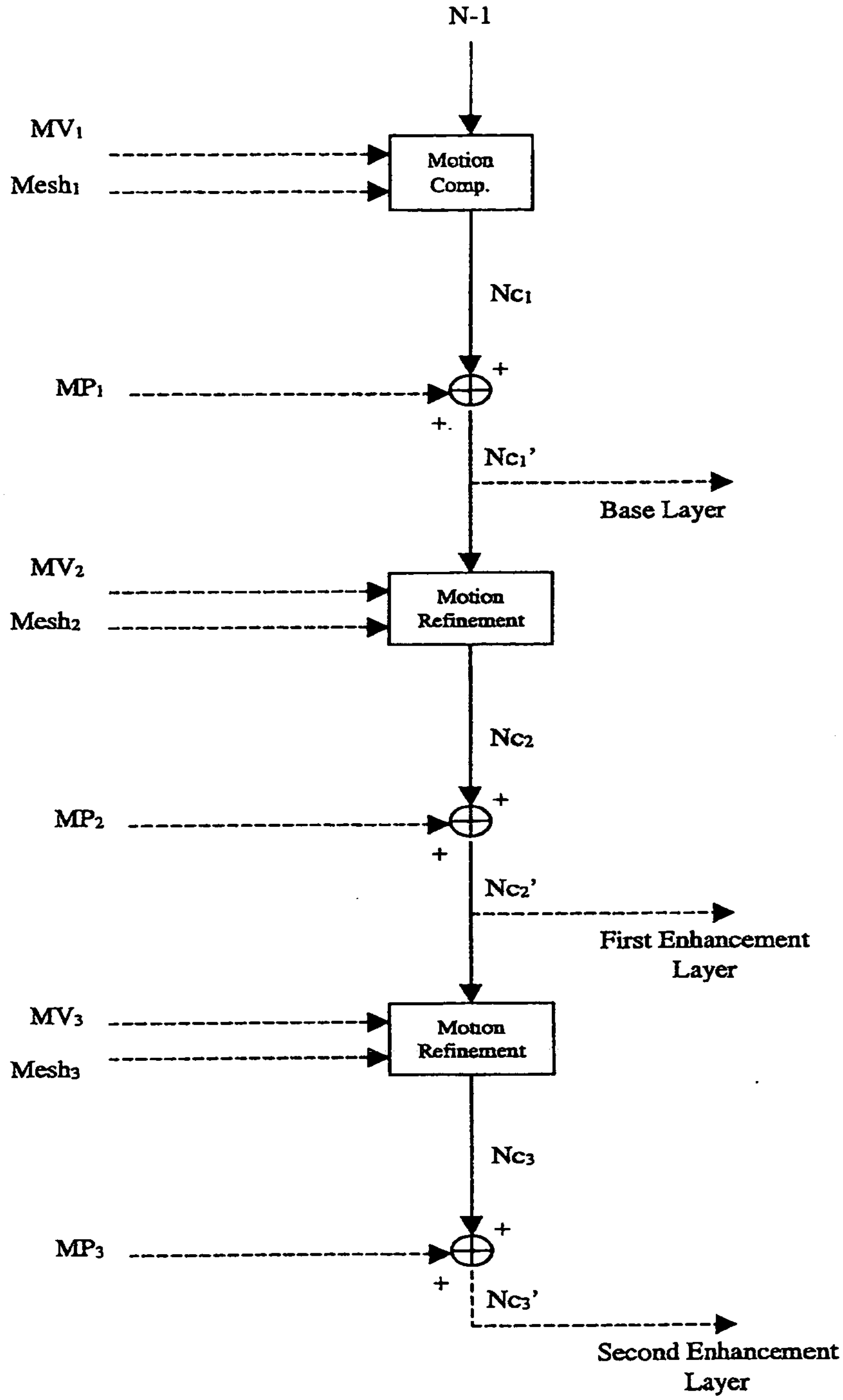

Figure 3: Block diagram of the decoder

This Page Blank (uspto)